A Comparative Evaluation on Data Transformation Approach for Artificial Speech Detection

ABSTRACT

The rise of voice biometrics has transformed user authentication and offered enhanced security and convenience while phasing out less secure methods. Despite these advancements, Automatic Speaker Verification (ASV) systems remain vulnerable to spoofing, particularly with artificial speech generated swiftly using advanced speech synthesis and voice conversion algorithms. A recent data transformation technique achieved an impressive Equal Error Rate (EER) of 1.42% on the ASVspoof 2019 Logical Access Dataset. While this approach predominantly relies on Support Vector Machine (SVM) as the backend classifier for artificial speech detection, it is vital to explore a broader range of classifiers to enhance resilience. This paper addresses this research gap by systematically assessing classifier efficacy in artificial speech detection. The objectives are twofold: first, to evaluate various classifiers, not limited to SVM, and identify those best suited for artificial speech detection; second, to compare this approach's performance with existing methods. The evaluation demonstrated SVM-Polynomial as the top-performing classifier, surpassing the end-to-end learning approach. This work contributes to a deeper understanding of classifier efficacy and equips researchers and practitioners with a diversified toolkit for building robust ASV spoofing detection systems.